

10/608, 083 PD-892

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 May 2002 (23.05.2002)

PCT

(10) International Publication Number
WO 02/41185 A1

(51) International Patent Classification⁷: **G06F 17/30** (74) Agent: ONSAGERS AS; P.O. Box 265 Sentrum, N-0103 Oslo (NO).

(21) International Application Number: PCT/NO01/00460

(22) International Filing Date:
20 November 2001 (20.11.2001)

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2000 5867 20 November 2000 (20.11.2000) NO

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

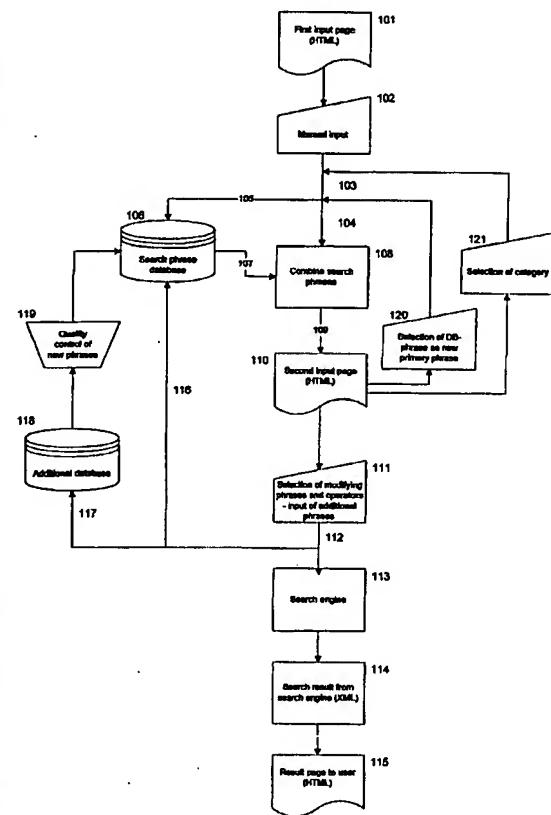
(71) Applicant (*for all designated States except US*): A/S E. PEDERSEN & SØNN [NO/NO]; Theresesgt. 28, N-0168 Oslo (NO).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): BREIVIK, Jarle [NO/NO]; Åsensvingen 5C, N-0488 Oslo (NO).

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR INTERACTIVE BUILDING AND OPTIMIZATION OF SEARCH EXPRESSIONS



(57) Abstract: A method and a system for building and optimizing search expressions is presented. The method is based on the reception of a primary search phrase, a search through a database for related search phrases along with associated Boolean operators, a presentation of these related search phrases along with the Boolean operators to a user, a reception of information regarding which modifying phrases the user has selected and the generation of a completed search expression to be forwarded to a search engine. The method also includes steps for registering the frequency with which search phrases and associated Boolean operators are used, as well as steps for registering new search phrases not already present in the database.

WO 02/41185 A1



Declaration under Rule 4.17:

- *of inventorship (Rule 4.17(iv)) for US only*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

Method and system for interactive building and optimization of search expressions

5 The invention relates to a method for interactively generating a search expression in a system acting as a front end to a search engine, and a computer system device performing the method.

10 The amount of information available through interactive (two-way) communication networks like the Internet is growing rapidly, and the task of identifying particular items of information is becoming increasingly more difficult. It has become common for on-line users to utilize search engines to search the Internet for desired information, and many web sites permit users to perform searches to identify a small number of relevant items among a much larger domain of items. As an example, several web sites permit users to search for particular web sites in indexes of known web sites, e.g. Fast™, Google™. Similarly, many on-line merchants, such as 15 Amazon™, permit users to search for particular products among all of the products that can be purchased from the merchant. Other on-line services, such as PubMed (National Library of Medicine) and Britannica™, allow users to search for various articles and documents.

20 In order to perform a search, a user submits a query containing one or more query terms. A query server program of the search engine processes the query to identify any items that match the terms of the query. The set of items identified by the query server program is referred to as a «query result», which is typically presented to the user as a set of hyper-textual listing (hyperlinks) of the located items.

25 If the scope of the search is large, the result of a simple query may result in hundreds, thousands or even millions of hits. In order to find a single item or a small set of items, the user is therefore required to build more complex queries involving numerous words and phrases, linked by default or specified Boolean operators (AND / NOT / OR). The quality of the query result therefore relies heavily on the user's knowledge in the field of interest, as well as his or her understanding of Boolean 30 logic. To users searching for an unfamiliar topic, or users unfamiliar with Boolean logic, searching the Internet is therefore a time-consuming and often futile process.

35 Several search engines, such as Yahoo™, DirectHit™, and AltaVista™ attempt to assist the query building process by suggesting related query terms to the user as a part of a «search refinement» process. This allows the user to further refine the query and narrow the query result by selecting one or more related query terms that more accurately reflect the user's intended request.

There are several ways to generate a database of related query terms that can assist the user in the building of a more comprehensive query expression:

1. Related query terms may be generated by the search engine, using the contents of the query result, such as by identifying the most frequently used terms within the located documents.
- 5 2. WO 99/66427 by Bowman et al. describes how query term correlation data may be extracted from the search queries submitted by a plurality of users.
3. Related query terms may be copied or extracted from existing dictionaries, thesauruses, indexes of textbooks, and other written material.
- 10 4. Lists of related query terms may be generated by individuals or committees, with experience in particular areas, e.g. a specific scientific discipline.

Irrespective of how it is generated, the information in the database of related query terms may be presented to the user in a manner that facilitates the building of more comprehensive search expressions. For example, a user submits a query on the term «CELL», which returns a query result of several thousand items. In order to assist the user in the refinement process, query terms related to «CELL» are looked up in the database and presented the user (e.g. the terms «ANIMAL», «FUEL», and «PHONE»). The user is then invited to limit the query result, e.g. through an interface of hyperlinks, check boxes, or drop down menus, and the selected terms are added to the query, which returns a more restricted query result.

But although often helpful, the existing technology for assisting users in refining search queries have major limitations, particularly related to the use of Boolean logic. Most importantly, none of the existing systems can assist the user in which Boolean operators should be used in relation to specific query terms. The need for such guidance in Boolean logic is particularly evident in relation to the use of the word «or», which has nearly opposite meanings in everyday language and as a Boolean operator. E.g. the phrase «You will get fish or meat» implies that you will get one or the other. The query «FISH OR MEAT», on the other hand, returns all the items that contain the word «FISH», as well as, all the items that contain the word «MEAT», including items that contain both the word «FISH» and the word «MEAT». As the use of Internet search engines has become an everyday activity, also to people untrained in Boolean logic, this confusion between the exclusive and the inclusive «or» has grown to become a major problem. The current method for dealing with the problem is limited to hyperlinks to separate «help-page», where the use of Boolean operators is explained in a traditional textbook-like manner. There is no direct relationship to the query refinement process, and there is no assistance on which Boolean operator should be related to which query term.

In relation to the «or-problem», the existing systems for assisting query refinement are solely designed to remove unwanted items from the primary query result. Information of interest to the user, not comprised by the primary query can therefore not be picked up by these procedures. E.g. a search refinement process that start with the term «PETROL» will not be able to identify pages containing the synonyms «FUEL» or «GASOLINE», which could be of equal interest to the user. A user must therefore repeat the search procedure for a number of synonyms, and this task is further complicated by the fact that the synonyms must be provided by a user, whom might be totally new to the subject in question. In order to overcome these limitations, the search refinement process must allow, and assist, the use of the «OR»-operator, and this problem is not tackled by any present technology.

Similarly, the best method for limiting a query result to a convenient set of relevant items often involve combination of terms that should, and should not, occur in the document. This implies queries where different terms are related to either the «AND» or the «NOT»-operator. The use of these operators are generally more intuitive, and can easily be assigned to specific terms by putting + or - signs in front of them. However, there is no existing technology that can assist the user in the process of selecting which operator should go with which term, or that can suggest purposeful combinations of terms and operators.

The present invention addresses these and other problems by providing a search refinement system and method for building query expressions comprising different Boolean operators from a database of related query terms, and a system and method for generating such a database. The invention relates to a database of primary query terms correlated to a set of related query terms with designated Boolean operators.

On submitting a primary query term, the user is presented the set of related query terms in a context that indicates and clarifies their assigned Boolean operator. This interface allows the user to select the combinations of terms and operators which should be included in the final query expression, and may also allow the user to switch between operators, or add additional terms with assigned operators.

The database of related query terms may be generated by one of or by different combinations of the methods listed above. Uniquely to the present invention, related terms are assigned specific Boolean operators, reflecting their most appropriate function related to the primary query term.

In accordance with one aspect of the invention, each related query term has a counter for each of the three Boolean operators (AND / NOT / OR), and in one embodiment, these counters are updated automatically from the user constructed query expressions submitted to the search engine. Thus, the entire community of users can function as a committee, which by the building of their own query expressions «vote» for the most

appropriate combinations of terms and operators, and thereby share their own judgment and knowledge with the other users. Similarly, new combinations of related terms and operators contributed by the user during the query building process would contribute to the expansion of the database. In a preferred embodiment, the building 5 of the database is accomplished through a combination of automatic updating of counters and an editorial quality control of terms that are new to the database.

A dynamic user interface that communicates relationship between query terms and their assigned Boolean operator has a key function in the present invention. In accordance with one aspect of the invention, this involves interactive screen images 10 that combine textual and graphic information, preferably composed from standard computer protocol, such as HTML, Java, or ASP. In one preferred embodiment, the user builds query expressions by marking checkboxes related to specific combinations of terms and operators. In one preferred feature of the interface involves a two-step procedure by which the user first is invited to expand the range of the query by 15 selecting synonym terms with the OR operator, and second to limit (focus) the query by selecting terms with the AND or NOT operator. This will avoid the previously described problem of the «or-problem». To further assist the user, particularly useful combinations of terms and operator may be checked as default.

In another preferred embodiment of the invention, the related query terms are 20 presented to the user as hyperlinks to pages where the specific related term represents the primary term with a corresponding set of related terms and assigned Boolean operators. This feature links the terms in the database together in a web of related terms in which the user can navigate to find the most appropriate primary term.

The particular features of the invention are given in the attached independent claims, 25 with the dependent claims describing additional features and embodiments.

The invention will now be described in the form of an exemplary embodiment, with reference to the attached drawings.

- Fig. 1 Shows a block diagram of a possible embodiment of a computer system operating in accordance with the invention.
- 30 Fig. 2 Shows the structure of the different tables in the search phrase database according to an embodiment of the invention.
- Fig. 3. Shows an example illustrating entries in the search phrase database.
- Fig. 4 Shows a flow chart illustrating an embodiment of the method according to the invention.
- 35 Fig. 5. Shows a suggested user interface for a user entering a search phrase.

Fig. 6 Shows a suggested user interface where a user is invited to select modifying phrases and Boolean operators.

Figure 1 shows a block diagram illustrating a computer system 10 operating in accordance with the invention. The system comprises at least one processor unit 11 capable of performing searches in a search phrase database 12. The data stored in the search phrase database 12 is accessible to the processor through input/output means 13. The processor unit 11 preferably operates in accordance with instructions in the form of computer program software stored on a local storage device 14, such as a hard drive. The processor unit 11 also has access to system memory 15 for temporary loading and unloading of instructions and data. Finally the system 10 comprises input/output means 16 for receiving search phrases from a user computer 20 and for sending a presentation of suggested modifying phrases and Boolean operators to the user computer 20. The system 10 also forwards completed search expressions to a search engine 30 over I/O-means 16, receives search results, and forwards these search results to the user computer 20. The communication between the computer system 10, the user computer 20 and the search engine 30 will normally take place over a public communication network 35 such as the Internet.

In a preferred embodiment the computer system 10 also comprises an additional database 17. This database is used for storing search phrases that are not entered into the main database 12 when such phrases are entered by a user, and to register the number of times these phrases occur. The data in the additional database 17 may then be used in order to refine the contents of the main database 12, either as a result of statistical analysis of the data collected in the additional database 17 over time, or based on evaluation of the data performed by humans.

A person skilled in the art will understand that the system 10 may be constructed in a number of different ways. It is possible to include all the units described in one single computer, or they may be distributed among several different computers. As an example, the two databases 12, 17 may be installed on separate computers. If this is the case, the two I/O-units 13, 16 may be one and the same. If, on the other hand, the databases 12, 17 are installed on the local storage unit 14, the I/O-means 16 will be part of the internal structure of the computer including the processor 11 and communication between the different units will mainly take place on this single systems system bus. Alternatively, the system may consist of several processors 11 in one or several computers, such as a multiprocessor computer or a cluster. Any such particular configuration is available to a skilled man and it will be a question of design choice which configuration is found preferable when building a system according to the invention.

Turning now to figure 2, the structure of records and tables in a search phrase database according to an embodiment of the invention is illustrated. The database comprises at least two different tables, the phrase table and the phrase relation table. Preferably there is also a third table called the category table. Preferably, the phrase 5 table contains all phrases in the database. Alternatively this table could contain only main phrases, i.e. phrases that are listed as primary search phrases, but for convenience and speed it has been found preferable to include all phrases in this table.

In the phrase table, every phrase is entered with one record for every category it appears in. For every relation between a main phrase and a modifying phrase, there is 10 a record in the phrase relation table. The phrase relation table also includes the counters that indicate how often a phrase is used as a modifying phrase.

The different fields of the records will now be described, starting with the phrase table.

The first field is *phrase_id* which is a key identifying the individual records of the 15 phrase table. *category_id* is a field that refers to a record in the *category* table. This indicates a field of search, such as a field within science, and will be used to make the suggested modifying phrases more targeted. This is explained in further detail below. It should be noted that while present in a preferred embodiment of the invention, the category table, and hence the *category_id* field of the records in the phrase table, may 20 be omitted.

The following field is the *phrase* which is a character string containing the text of the actual phrase. As already mentioned, according to a preferred embodiment all 25 phrases, whether they are primary phrases or only modifying phrases, are listed in the phrase table. In order to immediately identify a phrase as either a primary phrase or a modifying phrase, there is preferably a field called *main_phrase* in the records of the phrase table, and this field simply indicates whether this phrase is only a modifying phrase, or also a main phrase. E.g. if the phrase «cell» is present both as a main phrase, or primary phrase, but also as a modifying phrase for the phrase «phone», this field will have the value YES. On the other hand, if the phrase «fast» has no entry as a 30 primary phrase, only as a modifying phrase, e.g. for the primary phrase «food», the value of this field will be NO.

Finally, the field *search_count* is a counter that is incremented once every time the phrase is actually included in a completed search expression. Alternatively it is only 35 incremented every time the phrase is used as a primary phrase, since modifying phrases are counted in the phrase relation table, as described below.

In a preferred embodiment, the database comprises a category table, which functions to narrow down the search by making it possible to distinguish between identical

words that have different meaning within different fields of knowledge. As an example, one category may be biology, another may be telecommunications, and the word cell may then have two separate entries in the phrase database, each referring to a different record in the category database. The category records only contain the 5 fields *category_id* and *category_name*. The first is the key identifying the category record; the second is the text string containing the name of the category.

The records of the phrase relation table will now be described. Every record in this table defines a relation between a main phrase and a related phrase (or primary phrase and modifying phrase). The first field is the *phrase_relation_id*, which is a key 10 identifying the individual records of this table. Following this is the *search_phrase_id*, and then the *related_phrase_id*. In the embodiment described above, where all phrases are present in the phrase table, whether they are main phrases or only modifying phrases, the *search_phrase_id* field refers to the *phrase_id* of a phrase that is listed as main phrase in the phrase table. E.g. if the phrase «cell» is present as a 15 main phrase and has modifying phrases related to it, there will be one record in the search relation table for each such related phrase. The field *search_phrase_id* will then refer to the record for the phrase «cell» in the phrase table in each such entry in the phrase relation table. The *related_phrase_id* refers to a phrase modifying the search phrase. As an example, if the *search_phrase_id* refers to the phrase «cell», the 20 *related_phrase_id* may refer to the record for the phrase «cancer» in the phrase table.

Normally a search phrase has several phrases listed as modifying phrases, and as already explained there will be one entry in the phrase relation table for each such modifying phrase. Using the example above, the phrase «cell» may have the phrases «cancer» and «phone» as modifying phrases. In this case there will be two entries in 25 the phrase relation table. Both will refer to the entry for «cell» in the phrase table in their *search_phrase_id* fields, but in their *related_phrase_id* fields they will refer to the entries in the phrase table for the phrases «cancer» and «phone» respectively.

This example does not take category into account. If the database does include a category table, however, a given phrase will be entered once in the phrase table for 30 each category in which it occurs. In this is the case, one record for the phrase «cell» in the phrase relation table will in its *search_phrase_id* field reference the phrase «cell» of the category «telecommunication» and in its *related_phrase_id* field it will reference the phrase «phone» also of the category «telecommunications». Another record in the phrase relation table will in its *search_phrase_id* field reference the 35 phrase «cell» of the category «biology» and in its *related_phrase_id* field it will reference the phrase «cancer» also of the category «biology».

In general it may be stated that each record in the phrase relation table will refer to one and only one search phrase to be modified (in the *search_phrase_id* field) and

one and only one modifying phrase (in the *related_phrase_id* field), but that each record of the phrase table may be referred to by any number of records in the phrase relation table, both as main search phrase and as modifying search phrase.

Following these are three fields called *counter_plus*, *counter_minus*, and *counter_or*.
5 These counters are incremented every time the modifying phrase is used with the Boolean operators AND, NOT and OR, respectively. Alternatively there is only one counter indicating how many times the modifying phrase has been used, but not taking into account with which Boolean operator it has been used.

Finally there is a field called *basic*. This field indicates whether the phrase should be
10 used as a modifying phrase with the given primary phrase if the user has selected a basic search rather than an interactive search. This will be described in detail below.

Figure 3 shows an additional example of entries into the database. Four records of the phrase table are shown, the entries for the phrases «inbreeding», «interbreeding», «product» and «products». These all have unique *phrase_id* and they all belong to the same category. Only «inbreeding» can be a main phrase, the other phrases can only occur as modifying phrases, as indicated by the values of the *main_phrase* field. The phrase relation table defines the relationship between these phrases. There is one entry in the phrase relation table for each relationship. In each such entry the *phrase_id* for the search phrase «inbreeding» is identified. Further, the *phrase_id* for
15 the respective modifying phrase is identified in the field *related_phrase_id*. In this way the record defines a link between a main phrase and a modifying phrase. The field *related_phrase* contains the text string for the relevant modifying phrase and duplicates the text string in the *phrase* field of the record in the phrase table referenced by the *related_phrase_id*. The counter fields show that the phrase «interbreeding» has been used 1000 times with the Boolean operator AND, while «product» and «products» have been used 1000 times each with the Boolean operator
20 NOT in search expressions with «inbreeding» as the primary phrase. None of the modifying phrases have been used with any other Boolean operators.

The field *basic* indicates that if the user chooses to perform a basic search with
30 «inbreeding» as the primary phrase, «interbreeding» should be included in the search expression, while «product» and «products» should not. Alternatively this field could also indicate which Boolean operator the phrase should be included with, or that it should not be included at all, e.g. by using the character +, -, 1 and 0 for AND, NOT, OR and ‘not included’, respectively.

35 The additional database 17 has the same structure as the search database 12. Its function will be explained below.

Reference is now made to figure 4, where a flow chart illustrates the preferred embodiment of the invention. In a first step 101 a primary input page generated by the system 10 is presented on a user's computer. An embodiment of this primary input page is illustrated in figure 4, which will be described below. The user inputs 102 his primary search phrase, and possible additional phrases. The primary search phrase is then sent 105 to the search phrase database 12 and a search 106 is made for a record in the phrase table where the field *phrase*, and in a preferred embodiment also the category referenced by the field *category_id*, corresponds to the input from the user. If the phrase is not found, or if the field *main_phrase* indicates that this phrase is only listed as a modifying phrase in the database, no modifying phrases are returned 107 from the database, only the original primary phrase. Otherwise, a search is made through the phrase relation table for records referencing the primary phrase in their *related_phrase* field. All phrases that are found this way are returned 107 along with the primary phrase. In a preferred embodiment the phrases are returned with a suggested Boolean operator. The suggested Boolean operator will be the operator with the highest number of occurrences according to the counter fields. According to a preferred embodiment of the invention two operators may be suggested, one being OR, the other being AND or NOT. This is for purposes of generating a user friendly interface, as described below with reference to figure 5.

20 The phrases returned 107 from the database are combined with the additional user phrases 104 in a following step 108. Based on this a second input page is generated and presented to the user 110. An embodiment of this input page is illustrated in figure 5. After the user has selected which modifying phrases he or she whishes to include in the search expression and with which Boolean operators, a completed search expression is generated 112 and sent to the search engine 30. The search engine performs the search 113, the results are returned 114 to the computer system 10, and a result presentation is generated and sent 115 to the user computer 20, preferably as HTML.

25 The completed search expression is also used to update the database 12 and the additional database 17 in the following manner. In a first step 116 the search database is updated by incrementing the above described counters as follows. For the primary phrase, and optionally for all the modifying phrases, the field *search_count* is incremented by one. For the modifying phrases, the counter associated with the Boolean operator chosen in the completed search expression is incremented by one.

30 35 If the completed search expression included primary and/or secondary phrases that were not originally found in the search phrase database 12, these are sent 117 to the additional database. This database is updated 118 in a following step. For any phrase not already present in this database, a new record is created in the phrase table. If the phrase was used as a modifying phrase and there is no record already present in the

phrase relation table, a record is created in this table, the *search_phrase_id* is set to reference the same phrase in the phrase table, and the *related_phrase_id* is set to reference the primary phrase of the completed search expression. Then the counters are updated in the same way as for the search database 12.

5 With regular intervals, the data accumulated in the additional database 17 will be analyzed 119 for possible inclusion in the search phrase database 12. This may be done automatically by statistical analysis, or manually e.g. by an editorial board. As a result, the search phrase database 12 will be updated to include search phrases and phrase relations actually used by users, making sure the suggestions made by the
10 system according to the invention take advantage of the experience of its users.

In a preferred embodiment of the invention the user may, when presented with the second input page, choose to use one of the suggested modifying phrases as primary phrase if the phrase is listed in the database as a main phrase. Preferably this will be indicated by presenting the phrase as a hyperlink. All the user has to do then is to
15 click on the phrase, and a new search will be initiated 120 with the relevant phrase as primary phrase.

If, when presented with the second input page, the user chooses to change the category, a new search will be initiated 121 with the same primary phrase, but with a different category.

20 Figure 5 illustrates a suggested user interface of an input page 50 that will be displayed on the user computer 20 and that invites the user to enter a primary search phrase 51 and possible additional phrases 52, choose a category 53 and choose either basic search 54 or interactive search 55. If the user chooses interactive search, the primary phrase, along with any additional phrases and the category, are sent to the
25 front end computer system 10. There the method discussed above will be performed and a second input page will be generated and returned to the user computer 20. If the user chooses basic search, no second input window will be generated. The modifying phrases indicated as basic, as described above, will be included in the search expression with their default Boolean operators. The input page 50 will
30 preferably be displayed as a HTML page in a web browser running on the user computer 20.

Figure 6 illustrates a suggested user interface 60 that will be generated by the computer system 10 and sent to the user computer 20, inviting the user to refine his or her search. In a first field 61, phrases that will expand the search, e.g. phrases with OR as their suggested Boolean operator, will be presented. In a preferred embodiment, the phrases that are included in basic searches are presented as already
35

selected 62, while other phrases are presented but not selected 63. The user is of course free to select or deselect any of the phrases.

In a second field 64, phrases that narrow the search will be presented. These are phrases that have AND or NOT as their suggested Boolean operator. The phrases will 5 preferably be presented with either AND (+) 65 or NOT (-) 66 already selected, based on the highest counter value in the fields *counter_plus* and *counter_minus* described above. Alternatively only the phrases included in basic searches are preselected. It would still be possible to indicate which choice is suggested, e.g. by highlighting the alternative with the highest counter value 67. Phrases that also occur as main phrases 10 are preferably presented as hyperlinks 68. The user can then initiate a new search within the same category by clicking on the hyperlinked word 68. The user may also change the category 69.

If the user clicks on a hyperlinked word 68, only the relevant phrase and the category need be returned to the computer system 10. Likewise, if the user changes category 15 69, the new category will be returned along with the original primary phrase. Both these alternatives initiates a new look up in the search phrase database 12 and the generation of a new second input page. However, if the user submits the contents of the entire input page to the computer system 10 where a search expression is generated based on the selections made. The input window preferably includes fields 20 70, 71 where the user may include additional expanding 70 and limiting 71 phrases. The input window also includes a button 72 or some other means for the user to select in order to submit the selections to the computer system 10.

The selections illustrated in figure 6, with «bacterial» selected as an expanding phrase, «cytoplasma» entered as an additional expanding phrase, «cancer» and «brain» 25 selected as limiting phrases with the Boolean operator AND, «junk food» selected as a limiting phrase with the Boolean operator NOT, and «human» entered as an additional limiting phrase with the Boolean operator AND, will result in the following completed expression being generated by the computer system 10:

(cell OR bacterial OR cytoplasma) AND cancer AND brain AND human NOT junk 30 food

The invention has been described by way of an example where phrases and phrase relations are ordered in separate tables of a database. However, it must be understood that within the scope of the invention it would be possible to create a system with a different organization of data while still operating in accordance with the invented 35 method. The example must therefore be understood not in a limiting sense, whereas the scope of the invention is defined by the attached claims.

CLAIMS

1. Method for interactively generating a search expression in a computer system acting as a front end to a search engine,
characterized in the steps of
 - 5 - receiving a primary search phrase and optional secondary search phrases as input,
 - looking up in a search phrase database any modifying phrase listed as related to the primary search phrase,
 - returning at least some such modifying phrases along with the primary search phrase and combining them with any secondary search phrases,
- 10 - generating a first output message including all such modifying phrases, along with the primary search phrase and any secondary search phrases,
- transmitting said message including a request for a second input message indicating which modifying phrases should be included in a completed search expression,
- receiving information indicating which modifying phrases have been selected for inclusion in the completed search expression; and
- 15 - sending the completed search expression to the search engine.

2. Method according to claim 1,
characterized in that said modifying phrases are included in the first output message along with suggested Boolean operators and that any selected modifying phrase is included in the completed search expression with the suggested Boolean operator.
- 20
3. Method according to claim 2,
characterized in that the inclusion of suggested Boolean operators in the first output message also includes alternative Boolean operators that may be substituted for the suggested Boolean operators in the completed search expression.
- 25
4. Method according to claim 1,
characterized in that the completed search expression is returned to the search phrase database where respective counters are incremented for the primary search phrase and all modifying phrases, to the extent said phrases are included in the database.
- 30
5. Method according to claim 3,
characterized in that the completed search expression is returned to the search phrase database where respective counters are incremented for the primary search phrase and all modifying phrases in combination with the Boolean operator chosen, to the extent said phrases are included in the database.
- 35

6. Method according to claim 1,
characterized in that any phrase that is part of the completed search expression and not included in the search phrase database is sent to a second database where
5 - if not already present, a record is created for said phrase, and
- a counter is incremented for the phrase.

7. Method according to claim 6,
characterized in that the record created in the second database indicates whether said phrase was a primary phrase or a modifying phrase, and in the
10 case that the phrase was a modifying phrase said record also indicates the primary phrase with which it is related.

8. Method according to one of the previous claims,
characterized in that all primary phrases in the search phrase database belong to a category, and that any occurrence of the same phrase in more than one
15 category are treated as individual phrases, and that any primary search phrase received from a user also indicates one such category.

9. Computer system for, based on user input and user feedback, generating search expressions to be forwarded to a search engine,
characterized in that the computer system (10) comprises a processor unit (11), a first database (12), a local storage unit (14) and input/output means (13, 16), said first database (12) containing a list of search phrases and a list of relationships between main search phrases and modifying search phrases, and said local storage unit (14) containing a set of instructions that enables the processor unit (11) to
20 - receive, over the input/output means (16), first input message in the form of a primary search phrase and optional secondary search phrases,
- look up in the search phrase database (12) any modifying phrase listed as related to the primary search phrase,
- generate a first output message including at least some such modifying phrases
25 along with the primary search phrase and any secondary search phrases,
- transmitting, over the input/output means (16), said message including a request for a second input message indicating which modifying phrases should be included in a completed search expression,
- receiving, over the input/output means (16), said second input message indicating which modifying phrases have been selected for inclusion in the completed search expression,
30 - generating a second output message with a completed search expression including said main search phrase, said secondary search phrases and said modifying search phrases, and

- sending, over the input/output means (16), the message containing the completed search expression to the search engine.

10. Computer system according to claim 9,

characterized in that the list in said first database containing relationships between main search phrases and modifying search phrases includes information on a Boolean operator associated with the modifying search phrase and that the modifying search phrases are accompanied with their associated Boolean operator when they are included in said first output message.

11. Computer system according to claim 10,

characterized in that instructions stored on the local storage unit (14) enable the processor unit (11) to generate said first output message in the form of a HTML document, that, when displayed on a user computer (20) will display the modifying search phrases along with said associated Boolean operators as well as alternative Boolean operators.

15. 12. Computer system according to claim 9,

characterized in that said list of relationships between main search phrases and modifying search phrases includes counter variables, and that instructions stored on said local storage unit (14) enable the processor unit to examine the completed search expression and increment counters for each modifying search phrase included in the completed search expression as related to the main search phrase of the completed search expression.

13. Computer system according to claim 11,

characterized in that there are one of said counter variables for each alternative Boolean operator, and that the counter updated is the counter associated with the Boolean operator with which the modifying search phrase was associated in the completed search expression.

14. Computer system according to claim 9,

characterized in that it includes an additional database (17) with the same format as the first database (12), and that the local storage unit (14) includes instructions that enable the processor unit (11) to identify any search phrase included in the completed search expression that is not included in the list of search phrases on the first database, to register any such phrase in the additional database (17), if not already registered there, and to increment a counter representing the number of times said phrase has been used in a completed search phrase.

35 15. Computer system according to one of the claims 9 to 14,

characterized in that said first database (12) comprises

- a phrase table containing one record for every phrase in the database, said record

comprising a first field containing a phrase identifier uniquely identifying the phrase, a second field containing a text string representation of the phrase, and a third field containing an indication on whether said phrase is listed as a main phrase or only as a modifying phrase, and

5 - a phrase relation table containing one record for every relation between a modifying phrase and a main phrase, said record comprising a first field containing a phrase relation identifier uniquely identifying the phrase relation, a second field containing a search phrase identifier referring to one record in the phrase table representing the phrase to be modified (main phrase), a third field containing a related phrase identifier referring to one record in the phrase table representing the modifying phrase, and at least one additional field representing a counter indicating how many times said relation has occurred in a completed search expression.

16. Computer system according to claim 15,
characterized in that said at least one additional field in the phrase
relation table is a fourth field representing a counter indicating how many times said
relation has been connected by an AND operator, a fifth field representing a counter
indicating how many times said relation has been connected by a NOT operator, and a
sixth field representing a counter indicating how many times said relation has been
connected by an OR operator.

20 17. Computer system according to claim 15 or 16,
characterized in that said first database (12) includes a category table
containing one record for each field of knowledge or theme to which a search phrase
may belong, said record comprising a first field defining a category identifier uniquely
identifying the category, and a second field containing a text string representing the
category, and
- that the records of the phrase table comprises a fourth field containing a category
identifier referring to one record in the category table representing the category to
which the phrase belongs.

18. Computer program product comprising program instructions for causing a
computer to perform the method of any of the claims 1 - 8.

19. Computer program product according to claim 18,
characterized in that it is embodied on a record medium.

20. Computer program product according to claim 18,
characterized in that it is stored in a computer memory.

35 21. Computer program product according to claim 18,
characterized in that it is carried on an electrical carrier signal.

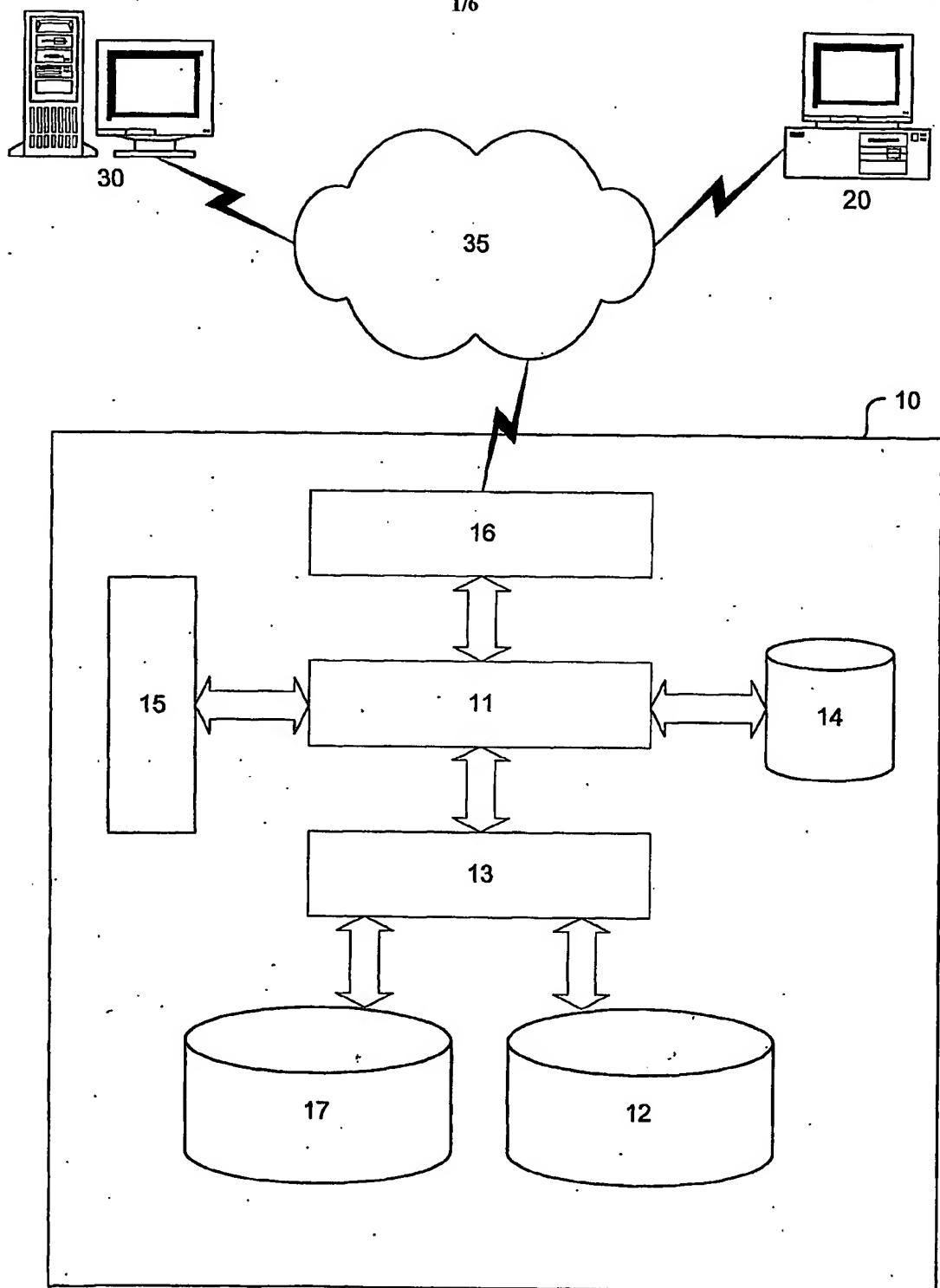
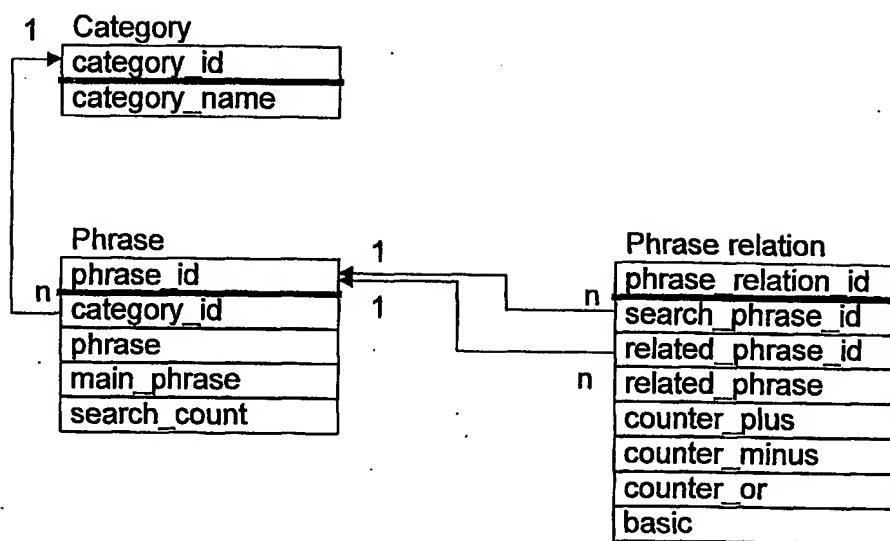


Fig. 1

*Fig. 2*

PHRASE table

<i>phrase_id</i>	<i>science_category_id</i>	<i>phrase</i>	<i>main_phrase</i>
55182	2	inbreeding	y
55183	2	interbreeding	n
55184	2	product	n
55185	2	products	n

PHRASE_RELATION table

<i>phrase_relation_id</i>	<i>search_phrase_id</i>	<i>related_phrase_id</i>	<i>related_phrase</i>	<i>counter_plus</i>	<i>counter_minus</i>	<i>counter_or</i>	<i>basic</i>
75839	55182	55183	interbreeding	1000	0	0	y
75840	55182	55184	product	0	1000	0	n
75841	55182	55185	products	0	1000	0	n

Fig. 3

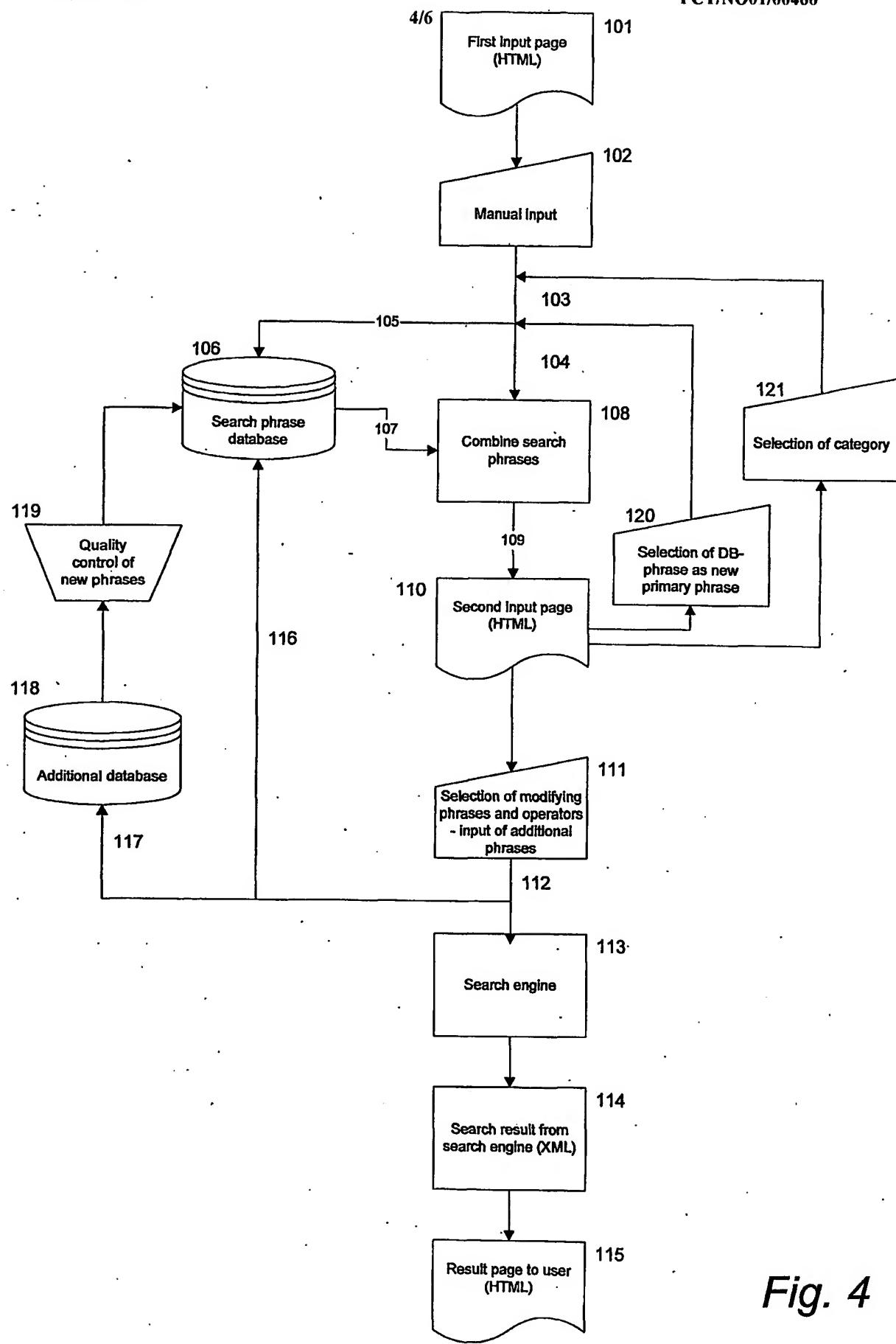


Fig. 4

50

Go directly to: <input type="text" value="Science Category"/>	
Search:	
Main word or phrases:	<input type="text" value="cell"/> 51
Additional Information:	<input type="text"/> 52
Select:	<input checked="" type="radio"/> Basic search <input type="radio"/> Advanced search The most simple More choices 54 55

Fig. 5

560

61 Category: <input checked="" type="checkbox"/> Biology <input checked="" type="checkbox"/>			Your main phrase: Cell
Expand your search These phrases are closely related to cell. Select them to get a more comprehensive coverage of the field.			
<input checked="" type="checkbox"/> 62 bacteria	<input type="checkbox"/>	<input type="checkbox"/>	70 ~
<input type="checkbox"/> 63 junk food	<input type="checkbox"/>	<input type="checkbox"/>	
Add your own synonyms or related phrases: <input type="checkbox"/> cytoplasm			
Limit your search These phrases are subdivisions of cell. Select them to get a more focused selection of the field.			
(+) 65 <input checked="" type="checkbox"/> cancer	(+) 66 <input type="checkbox"/> blood	(+) 67 <input type="checkbox"/> tumor	71
<input checked="" type="checkbox"/> 68 bacterial	<input type="checkbox"/> biology	<input type="checkbox"/> nerve	
<input checked="" type="checkbox"/> 69 brain	<input type="checkbox"/>	<input type="checkbox"/>	
Add your own limiting phrases: <input checked="" type="checkbox"/> human			
64 ~ 72 ~ <input type="button" value="Submit"/>			

Fig. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/NO 01/00460

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: G06F 17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 9966427 A1 (AMAZON.COM), 23 December 1999 (23.12.99), page 2, line 12 - page 3, line 15 -----	4-7,12-14

Further documents are listed in the continuation of Box C.

See patent family annex.

- * Special categories of cited documents
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 March 2002	20-03-2002
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Oskar Pihlgren / MRO Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/NO 01/00460**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: **1-3, 8-11, 15-21**
because they relate to subject matter not required to be searched by this Authority, namely:
see extra sheet
2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NO 01/00460

The subject matter claimed in claims 1-3, 8-11, 15-21 lacks technical features, and therefore these claims are not required to be searched.

In claims 1-3, 8-11, 18-21 is described, besides obvious and well known technical features such as a database a processor etc, only an administrative method and system for a user to choose among different search expressions, this is done with mental calculus and is not technical.

What is stated in claims 15-17 describes only the layout of a database structure. This is not required to be searched with reference to Rule 39.1 (iii).

INTERNATIONAL SEARCH REPORT

Information on patent family members

28/01/02

International application No.

PCT/NO 01/00460

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9966427 A1	23/12/99	AU	4678399 A	05/01/00
		EP	1104567 A	06/06/01
		US	6006225 A	21/12/99
		US	6085195 A	04/07/00
		US	6169986 B	02/01/01